FILE COPY FOREST INSECT LABORATORY
UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

460.c
45K

UNITED STATES DEPARTMENT OF AGRICULTURE

BUREAU OF ENTOMOLOGY AND PLANT QUARANTINE

Project	Date	Author

TITLE

VARIATION IN THE APPLICATION OF TREE CLASSIFICATION SYSTEMS TO PONDEROSA PINE

JMK
MLL
KAS
JEP
RCH
JSY
BTC
PCJ
ESK

By
C. B. Eaton
Berkeley, California
March 11, 1942

U. S. GOVERNMENT PRINTING OFFICE 8—0742

SUBJECT—

INDEX NO.—

Copies sent to —

Craighead — 2

Kern

Evenden

R. O.

Experiment Station

Forest Insect Laboratory
Berkeley, California
March 11, 1942

VARIATION IN THE APPLICATION OF TREE CLASSIFICATION SYSTEMS TO PONDEROSA PINE

APPROVED BY:

J. M. MILLER
Senior Entomologist, In Charge

SUBMITTED BY:

G. B. EATON
Assistant Entomologist

TABLE OF CONTENTS

IMPORTANCE OF UNIFORMITY - 1

FIELD PROCEDURE - 2

ANALYSIS OF THE PROBLEM - 3

CONSISTENCY IN TREE CLASSIFICATION - - - - - - - - - - - - - - 3

Ratings of the Same Man - - - - - - - - - - - - - - - - - - 3

Standard Ratings - 5

Distribution by Class - - - - - - - - - - - - - - - - - - - 5

Comparison of Systems - - - - - - - - - - - - - - - - - - - 8

CONSISTENCY IN APPRAISING TREE CHARACTERS - - - - - - - - - - 8

SUMMARY AND CONCLUSIONS - - - - - - - - - - - - - - - - - - - 10

LITERATURE CITED - 11

VARIATION IN THE APPLICATION OF TREE CLASSIFICATION SYSTEMS TO PONDEROSA PINE

Within the last decade three tree classification systems for ponderosa pine (Pinus ponderosa Laws.) have been described by Dunning (1), Keen (3), and Salman and Bongberg (5) respectively. These systems have become increasingly important in the development and guidance of marking practices and research studies in western United States. Consistent use of all tree classifications thus far proposed depends chiefly upon the abilities of the observer, since classes are generally separated by characters determined by visual appraisal. The degree of variation from this source was made the object of a study by the Bureau of Entomology and Plant Quarantine at the Black's Mountain Experimental Forest of the California Forest and Range Experiment Station in 1940. An analysis of the results of this study is presented in this report.

IMPORTANCE OF UNIFORMITY

Among men who are confronted with the task of applying tree classification systems on both timber sales and research projects, the question that recurs perhaps most frequently is whether or not the individual can consistently rate the same trees alike and in agreement with other individuals of similar training and experience. Perhaps nothing is more exasperating to the average timber marker than the attitude of supervising officials who pounce upon one of perhaps many thousands of marked trees with the question, "Why did you mark that tree?" The ease with which one can distinguish tree classes becomes a matter of real importance, not so much for the peace of mind of the timber marker as for assuring uniformity in the results of cutting practices. In detailed research studies, where tree classes are used as standards by which other variables are to be measured, the fact is often overlooked that the variation in applying the classification may be so great as to seriously disturb the results of the experiment. For example, in a study set up to determine the change in the tree class distribution in a given stand over a period of years, it would be highly desirable to know how accurately ratings on individual trees could be applied in order to know whether or not the results secured from the study were due to real changes in the trees or changes in ideas of the observer.

In published literature no mention has been made of studies on consistency or uniformity in classifying trees. Dunning (1) observed that while the attainment of perfect agreement between different men was not to be expected, adherance to a system of appraisal based on easily discernible characters would improve the quality of the marking practice. In referring to the use of the Dunning system, Meyer (4) states that after a little practice with an experienced estimator, the determination of tree

class may be readily made at a glance. Keen (3) makes no reference to the identification of his tree classes in the field beyond implying that they are easily recognized. In describing a modification of Keen's tree classification system for Black Hills ponderosa pine, Hornibrook (2) stressed the fact that a tree classification is a useful but not foolproof tool to aid the inexperienced in becoming expert at the art of timber marking. Salman and Bongberg (5), in their discussion of the development of risk ratings for trees, state that the Risk system is well adapted to field use, and may be adequately applied by means of simple ocular examination. In referring to a classification for lodgepole pine (Pinus contorta latifolia Engelm.) Taylor (6) states that the mental image of each class becomes clearer with each trial, but he makes no reference to the ultimate degree of perfection attainable in recognizing individual classes.

FIELD PROCEDURE

On a plot located in an uncut stand of ponderosa pine, site quality III-IV, as shown in figure 1, 100 trees above 11.6 inches diameter at breast height were marked for observation. Trees above the minimum diameter limit were taken as they occurred in the stand, without discrimination. The average diameter of the group was 26 inches, with a range of 12 to 45 inches, and the average height was approximately 100 feet. All trees were numbered in order to identify each set of observations with the individual tree.

The observers or raters consisted of five men representing three levels of experience. For reporting purposes they have been designated "A", "B", "C", "D", and "E", in the order of their experience, proceeding from "A", the most skillful, to "E", the least. Specifically the background of the observers was as follows: "A", expert, with four seasons' work in tree classification; "B" and "C", experienced, with two seasons' work; and "D" and "E", inexperienced. In this test the expert and the experienced men rated the trees according to their previous training. The observations of the inexperienced men were based upon their interpretation of written or illustrated descriptions of the tree classes.

In the field each man was issued a series of forms on which were listed the tree classes together with the following characters used in describing trees: dominance, age, crown length, crown symmetry, crown width, top shape, appearance, crown density, needle length, needle color, twig injury, branch injury, and percent topkill. The list was prepared from the written descriptions of tree classification systems. Qualitative gradations of each character were provided for upon the form, and each observer was asked to check the grade of the character that seemed to him most applicable for the particular tree under consideration. For example, branch injury gradations were listed as none, light, moderate, or severe; crown density as dense, medium, or thin; etc. The tree class was then rated by Dunning's, Keen's, and the Risk systems. This procedure differs from the customary procedure in that in the present instance the marker made his rating after

considering the tree characters individually, whereas in ordinary marking practice the observer makes his appraisal at a glance. Two sets of ratings were made on the trees two weeks apart. This time interval was considered long enough to insure that the rater would forget the concept of any particular tree, but would be short enough so that there would be little or no physical changes in the trees. Each observer worked independently of the others, and to avoid bias in the two sets of observations, each used a separate set of forms on the two occasions.

ANALYSIS OF THE PROBLEM

In setting up the test, the expectancy was that the results should yield some measure of the uniformity and consistency with which trees are rated under different systems. Under this expectation answers were sought to the following questions: How often will the same man classify the same trees alike on different occasions? What is the maximum number of times a well trained marker will classify the same trees alike? Are differences in ratings on the same trees compensating, i.e., will a marker place the same number of trees in each class on successive occasions, even though he disagrees in individual cases? How well do ratings of different men of the same experience level compare? Can untrained men accurately classify trees aided only by written descriptions of the tree classification? Are all tree classification systems used with the same degree of proficiency? Are all crown characters of the same tree seen alike by different men, and if not, what crown characters are most readily recognized? These are practical questions, the answers to which should be useful to all who are engaged in studies in which tree classification systems are used, since they have a direct bearing on the significance of the results.

CONSISTENCY IN TREE CLASSIFICATION

In analyzing the data it was found that consistency or uniformity could be shown by the following comparisons: (1) frequency of agreement between two sets of ratings on the same trees by the same observer; (2) frequency of agreement with standard ratings by different observers; (3) tree class distribution by different observers; (4) frequency of agreement in the different tree classification systems.

Ratings of the same man.-- Frequency of agreement in two sets of ratings on the same trees by the same observer is shown for Dunning's, Keen's and the Risk classifications in Table 1 and in Figure 2. It will be seen from a study of this table that the best agreement in the two sets of observations was attained by the most experienced man, Rater "A", except with Dunning's classification where "B" made the highest score. With Dunning's system "B" reclassified 85 trees the same on the second occasion as on the first. With Keen's system "A" reclassified 70 trees alike, and with the Risk system 83 trees alike. The poorest agreement in the two sets of observations was attained by rater "E" who classified 28, 14, and 52 trees alike with Dunning's, Keen's, and the Risk systems respectively. The scores

Table 1.-- Frequency of Agreement of Individual Observers Classifying the Same Trees on Two Occasions.

Observer	Number of Trees Classified Alike			Number of Trees Classified Differently
	Dunning's Classification			
A	81			19
B	88			12
C	67			33
D	53			47
E	28			72
	Keen's Classification			
A	70	Age	87	13
		Vigor	81	19
B	56	Age	86	14
		Vigor	72	28
C	53	Age	73	27
		Vigor	63	37
D	38	Age	59	41
		Vigor	61	39
E	14	Age	20	80
		Vigor	58	42
	Risk Classification			
A	83			17
B	72			28
C	65			35
D	68			32
E	52			48

of "B", "C", and "D" fell between these extremes with the exception noted above. According to the laws of chance the probability that the ratings would match is 14.29 percent for Dunning's system, 6.25 percent for Keen's, and 25 percent for the Risk system.

Standard ratings.--Frequency of agreement with a standard set of ratings for Dunning's system and the Risk system is shown in Table 2.

Table 2.— Frequency of Agreement of Individual Ratings with Standard Ratings1/ on the Same Trees.

Observation	Observer A	B	C	D	E
		Dunning's Classification			
First	100	95	88	60	84
Second	81	75	75	66	24
		Risk Classification			
First	100	52	59	44	61
Second	83	64	53	58	60

1/ Ratings applied by the most experienced man assumed correct for purposes of comparison.

In this comparison agreement is expressed as a percentage of the ratings applied by the most experienced individual. The latter, for purposes of comparison, is assumed to be 100 percent correct. For example, the ratings applied by "B" with Dunning's system in the first set of observations were the same on identical trees 95 times out of 100 as those applied by "A". The frequency of agreement ranged from 24 to 95 percent for Dunning's system, and from 44 to 83 percent for the Risk system. Because of the lack of a set of ratings that could be considered sufficiently accurate, and the difficulty of making this kind of a comparison for a classification with a large number of classes, Keen's system was not analyzed in this manner.

Distribution by class.— An analysis in which the ratings were grouped by class was made to show whether or not disagreement in the ratings of specific trees is compensating (1) for two sets of ratings made by the same man, and (2) for ratings made by different men. These data are summarized in Table 3. It will be seen from this table that the number of trees placed in each class is compensating to a degree that brings about somewhat better results than is the case where ratings of individual trees are compared. For example, according to rater "A's" observations, 89 percent of the trees had identical distribution by classes with Dunning's system, but on individual trees "A's" ratings harmonized in only 51 cases out of 100. Harmony in the number of trees distributed among the various classes varied according to the experience of the observer, its absence being the most noticeable in the case of the inexperienced men.

Table 3.— Distribution of Tree Class Ratings Applied to the Same Trees By Different Men

Class	A		B		C		D		E	
	Observation									
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Dunning's Classification										
1	11	16	13	10	10	14	9	14	19	45
2	5	4	4	3	7	11	2	3	1	4
3	19	21	25	32	20	20	35	40	27	28
4	20	13	5	5	12	12	12	3	1	9
5	42	43	47	42	44	41	20	36	40	5
6	0	0	0	0	0	1	15	1	4	5
7	3	3	6	8	7	1	6	3	8	4
Keen's Classification										
1A	1	1	4	1	1	5	7	0	10	14
1B	0	0	0	0	0	0	0	0	0	6
1C	0	0	1	0	1	1	3	0	0	1
1D	0	0	0	0	0	0	1	0	1	0
2A	7	10	6	6	6	7	8	10	9	10
2B	7	8	5	5	6	7	2	5	0	15
2C	1	1	1	1	1	3	2	2	0	7
2D	0	0	0	0	0	0	2	0	0	1
3A	9	7	10	11	9	15	25	20	22	9
3B	21	21	14	18	14	11	21	17	5	21
3C	9	6	4	8	7	7	0	6	4	11
3D	1	1	1	2	2	2	2	1	2	0
4A	8	9	12	10	8	11	0	5	21	2
4B	22	22	22	17	20	16	17	25	16	2
4C	12	12	16	15	16	11	2	7	9	1
4D	2	2	4	6	9	4	7	2	1	0
Risk Classification										
1	34	38	58	45	39	50	46	53	32	50
2	38	39	29	38	34	22	28	25	29	20
3	13	11	10	13	11	13	7	8	19	10
4	15	12	3	4	16	15	19	14	20	20

It is of interest to note from the data in Table 3 that better agreement is secured when the ratings are grouped according to their distribution by classes, rather than by frequency of agreement on individual trees. In this analysis agreement is better with Keen's system than with the Risk or Dunning's systems. This fact can be substantiated by the use of the chi-square test (Table 4). In order to fully understand the

Table 4.-- Chi-square for Tree Class Distribution in Two Sets of Ratings by Different Observers

System	Observer									
	A		B		C		D		E	
	Chi	P.O.*	Chi	P.O.	Chi	P.O.	Chi	P.O.	Chi	P.O.
Dunning	4.11	50%	2.61	95%	7.62	30%	28.53	1%	64.10	1%
Keen	1.61	99	6.73	95	10.23	50	35.55	1	94.78	1
Risk	0.93	50	5.79	10	6.26	10	2.20	50	12.15	1

*P.O. is the probability of occurrence that the value of chi-square will be exceeded in indicated percentage of samples.

application of the chi-square test here it must be recalled that the present experiment is predicated on the assumption that no perceptible changes in the trees would occur in the two week period between observations. Consequently, if a significant difference is shown by the chi-square test, between the number of trees placed in a particular class by any observer, the conclusion that the observer is at fault is the logical one. This expectation is fulfilled by the data in the table, where according to "A", the expert, there is no significant difference between the number of trees placed in any particular class, but according to the inexperienced men "E" there is a significant difference under all systems. If the observations of "D" and "E" are not considered, it can be said that the likelihood that the differences in tree class distribution are due to chance is much greater for Keen's system than for Dunning's or the Risk system. Apparently disagreement on borderline trees is compensating, thereby bringing about better uniformity in the distribution of classes for the former system than for either of the latter.

Reference again to Table 3 shows that there is a wide range of difference between individual observers as to the number of trees that each distributed among the various classes. In nearly all classes the chi-square test indicated that significant differences existed between the maximum and minimum number of trees placed in the class by the several men. Here again, however, the differences are attributable to changes in the observers' interpretations and not to actual changes in the trees. The changes are not compensating when men of all three experience levels are compared; however, if the observations of the two inexperienced men, "D" and "E", are disregarded it will be seen that the results compare fairly well.

Comparison of systems.-- It would be expected, if for no other reason than the fact that the systems differ in number of classes, that variation would occur between Dunning's, Keen's and the Risk systems, i.e., that all systems would not be used with equal facility. In the present case this expectation held true. With experienced men the frequency of agreement in the two sets of observations of each man on specific trees was greater with Dunning's system than with the Risk system, and it was the least with Keen's system (Figure 2). Frequency of agreement with standard ratings was better with Dunning's system than with the Risk system (table 2). However, if harmony in the ratings on specific trees is not taken into account, thereby allowing differences to compensate for each other, it appears that Keen's system was more uniformly applied than either the Risk or Dunning's system.

The frequency of agreement of ratings in two sets of observations made on specific trees by different numbers of men is shown in Figure 3. For example, with Dunning's system at least one of the markers (but not necessarily the same man in every case) classified each of the 100 trees the same on both occasions. With Keen's system 92 of the trees were re-classified the same by at least one man, but the class of 8 trees was not agreed upon by any one of the observers. The number of trees upon which all five observers collectively agreed in regard to class was 19 with the Risk system, 7 with Dunning's and 2 with Keen's. The spiral curves that show the relationship between the number of men and the frequency of agreement, reflect the influence of the three experience levels among the observers.

CONSISTENCY IN APPRAISING TREE CHARACTERS

Since tree classification is based on the appraisal of tree characters, it was expected that an analysis of character ratings would offer some explanation for the discrepancies noted above in the application of the various systems. In Table 5 a comparison is given of agreement in tree

Table 5.--Frequency of Agreement of Individual Observers Appraising Tree Characters on the Same Trees on Two Different Occasions

Character	Observer				
	A	B	C	D	E
Crown length	88%	88%	84%	68%	80%
Top shape	84	80	80	68	76
Crown density	84	84	92	56	60
Crown width	72	68	72	88	64
Dominance	84	96	76	24	84
Branch injury	84	72	80	76	48
Appearance	72	52	84	72	76
Needle length	68	36	84	72	84
Age	92	72	92	52	24
Needle color	92	76	96	0	68
Twig injury	84	72	60	60	56
Crown symmetry	68	48	68	72	60

character ratings made by the same observer. The character most readily appraised by all observers was crown length, followed by top shape. The character most difficult to appraise was crown symmetry, and the next most difficult were age, needle color, and twig injury. Great variation in frequency of agreement occurred among the five observers; however, better uniformity on nearly all characters was attained by the more experienced men. Tree character evaluations of none of the observers were in agreement 100 percent of the time, the highest score being 92. On the other hand one instance occurred, in the case of observer "D", where ratings of a character (needle color) did not agree in a single instance.

Frequency of agreement on tree characters appraised by different numbers of men is given in Table 6. Top shape and crown length were

Table 6.-- Frequency of Agreement on Tree Character Ratings by Different Numbers of Men

Character	Agreement				No Agreement
	2 men	3 men	4 men	5 men	
Top shape	22	18	14	8	3
Crown length	24	19	10	7	1
Dominance	22	17	9	4	3
Branch injury	20	13	6	3	5
Age	25	14	6	2	0
Appearance	23	12	6	1	2
Crown density	23	17	8	1	2
Crown symmetry	22	12	5	1	3
Crown width	22	13	7	1	3
Needle color	23	16	7	0	2
Needle length	23	11	5	0	2
Twig injury	23	15	8	0	2

appraised the same by all five men 33 percent of the time. Needle color, needle length, and twig injury were not appraised the same by all five men in any case. On all characters at least two (not necessarily the same observers) men agreed 88 to 100 percent of the time, whereas all five men agreed 0 to 33 percent of the time. Differences in the experience of the observers probably accounted to a large extent for the wide discrepancies in character ratings.

SUMMARY AND CONCLUSIONS

A study was made of the uniformity with which markers applied Dunning's, Keen's, and the Risk tree classification systems for ponderosa pine. Five men representing three levels of experience marked 100 trees on consecutive occasions two weeks apart. Analyses of their ratings on these trees indicate the following:

1. The ability of a marker to classify trees uniformly varies with his experience. A well trained man can rate individual trees alike on consecutive occasions approximately 80 percent of the time with Dunning's and the Risk systems, and 70 percent of the time with Keen's system.

2. Ratings applied to the same trees by men of equal experience compare favorably, but untrained men cannot satisfactorily apply tree classes guided only by published descriptions of the systems.

3. On a case history basis trained men can use Dunning's classification more uniformly than the Risk classification, and the Risk classification more uniformly than Keen's classification. On a tree class distribution basis the order of proficiency is reversed.

4. Discrepancies in tree class ratings made by different men may be accounted for by the fact that the discernment of tree characters is not critical. Characters like crown length, that may be compared against a standard, are more consistently rated than characters like twig injury that cannot.

The results of this study imply that real changes in the class of a tree cannot be accurately detected in periodic appraisals because of the variation inherent in the observer. The study suggests that such changes may be more precisely detected by comparing the tree class distribution in the stand as a whole on successive occasions.

LITERATURE CITED

1. Dunning, D. 1928. A tree classification for the selection forests of the Sierra Nevada. Jour. Agr. Res. 36:755-771.

2. Hornibrook, E. M. 1939. A modified tree classification for use in growth studies and timber marking in Black Hills ponderosa pine. Jour. Forestry 37:483-488.

3. Keen, F. P. 1936. Relative susceptibility of ponderosa pines to bark-beetle attack. Jour. Forestry 34:919-927.

4. Meyer, W. H. 1934. Growth in selectively cut ponderosa pine forests of the Pacific Northwest. U. S. Dept. Agric. Tech. Bul. 407. 64 pp. illus.

5. Salman, K. A. and J. W. Bongberg. 1942. Logging high-risk trees to control insects in the pine stands of northeastern California. Jour. Forestry (In press)

6. Taylor, R. F. 1937. A tree classification for lodgepole pine in Colorado and Wyoming. Jour. Forestry 35:868-875.

Figure 1.-- Virgin ponderosa pine type on the Lassen National Forest, California. This scene typifies stand conditions under which tree classification systems are used. (Photo by P. C. Johnson)

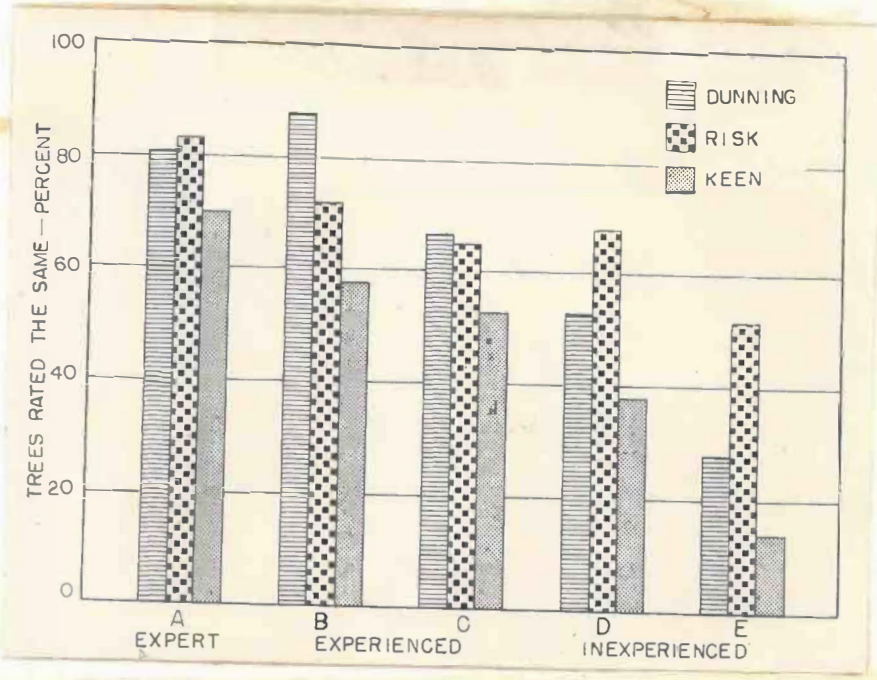

Figure 2.-- Frequency of agreement in two sets of tree class ratings made by men of different experience levels. Based on observations on 100 trees.

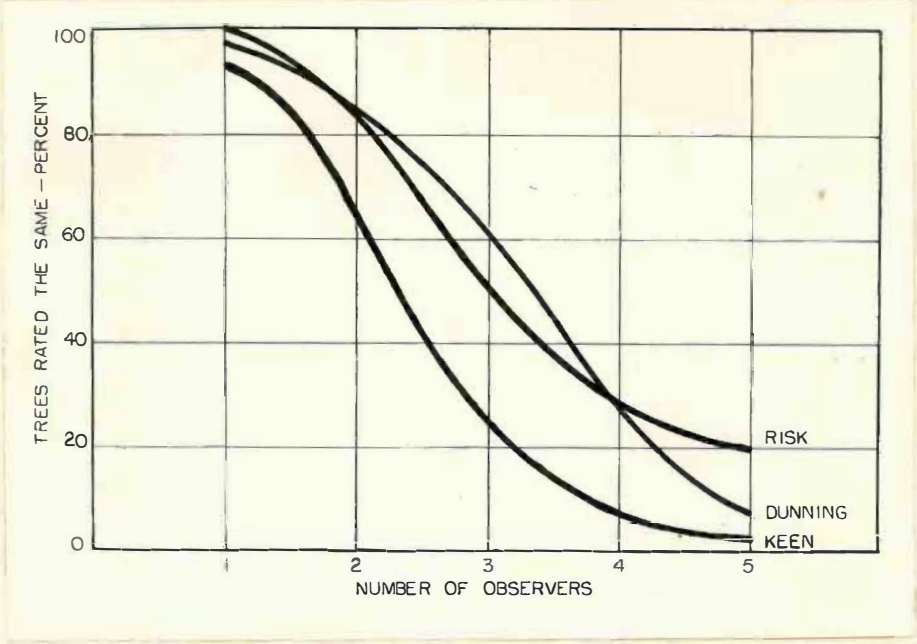

Figure 3.-- Frequency of agreement, on designated trees, of tree class ratings made by different numbers of men. Based on two sets of observations on 100 trees.